

Final Report
A Comprehension Based Analysis of Autoflight System
Interfaces

Peter G. Polson, Institute of Cognitive Science

University of Colorado

NASA Cooperative Agreement NCC 2-1104/Colorado

NASA Ames Research Center

Contact Monitor: Dr. Everett Palmer

Abstract
Final Report
A Comprehension Based Analysis of Autoflight System Abstract

This cooperative agreement supported Dr. Peter Polson's participation in two interrelated research programs. The first was the development of the Situation-Goal-Behavior (SGB) Model that is both a formal description of an avionics system's logic and behavior and a representation of a system that can be understood by avionics designers, pilots, and training developers. The second was the development of a usability inspection method (Nielsen and Mack, 1994) based on an approximate model, RAFIV, of pilot interactions with the Flight Management System (FMS).

The main purpose of this report is to integrate the two models and provide a context in order to better characterize the accomplishments of this research program. A major focus of both the previous and this Cooperative Agreement was the development of usability evaluation methods that can be effectively utilized during all phases of the design, development, and certification process of modern avionics systems. The current efforts to validate these methods have involved showing that they generate useful analyses of known operational and training problems with the current generation of avionics systems in modern commercial airliners.

This report is organized into seven sections. Following the overview, the second section describes the Goal-Situation-Behavior model and its applications. The next section summarizes the foundations of the RAFIV model and describes the model in some detail. The contents of both these sections are derived from previous reports referenced in footnotes. The fourth section integrates these two models into a complete design evaluation and training development framework. The fifth section contains conclusions and possible future directions for research. References are in Section 6. Section 7 contains the titles and abstracts of the papers describing in more detail the results of this research program.

**Final Report:
A Comprehension Based Analysis of Autoflight System
Interfaces**

**Peter G. Polson, Institute of Cognitive Science
University of Colorado**

1. Introduction and Overview

A major focus of both my previous and this Cooperative Agreement was the development of usability evaluation methods that could be effectively utilized during all phases of the design, development, and certification process of modern avionics systems.

This cooperative agreement supported Peter Polson's participation in two interrelated research programs. The first was the development of the Situation-Goal-Behavior (SGB) Model that is both a formal description of an avionics system's logic and behavior and a representation of a system that can be understood by avionics designers, pilots, and training developers (Sherry, Feary, Polson, and Palmer, 2000a, b; Sherry, Feary, Polson, Mumaw, and Palmer, 2001a; Sherry, Feary, Polson, and Palmer, 2001b). The SGB formalism is also the foundation of an automated avionics software development methodology that supports formal analyses of the completeness and consistency of a proposed design (Sherry, 1995). This activity was supported by the National Aeronautics and Space Administration (NASA) under contract NAS1-20219 to Honeywell Air Transport Systems (COTR: Everett Palmer) and cooperative agreements NCC 2-904 and NCC 1104 with the University of Colorado (COTR: Everett Palmer). Collaborators included Lance Sherry (then with Honeywell) and Everett Palmer and Michael Feary (NASA Ames).

The second was the development of a usability inspection method (Nielsen and Mack, 1994) based on an approximate model, RAFIV, of pilot interactions with the Flight Management System (FMS), (Sherry, Polson, Feary, and Palmer, 2002a; Sherry, Polson, Fennell, and Feary, 2002b). RAFIV is an acronym for the five stages in the model. The names of the stages are Reformulate, Access, Format, Insert, and Verify. This research was supported by the NASA Aviation Operations Systems Program contract GS09T01BHM0386 order ID 9T1N001MH to Honeywell International CAGR (TPC: Everett Palmer, Michael Feary) as well as Cooperative Agreement NCC 2-1104 with the University of Colorado. Collaborators included Lance Sherry (then with Honeywell), Everett Palmer and

Michael Feary (NASA Ames), Randall Mumaw (Boeing), and Karl Fennell (United Airlines).

1.1 Foundations

The theoretical and empirical foundations of the RAFIV model of pilot interactions with the avionics were developed in research program supported in part by a previous Cooperative Agreement, NCC 2-904. The details of the cognitive mechanism underlying RAFIV are based on a model developed by Kitajima and Polson (1997), **L**inked model of **C**omprehension-based **A**ction planning and **I**nstruction taking (LICAI). It is a comprehension-based, cognitive analysis of action planning and display comprehension. This work was done in collaboration with Dr. Muneo Kitajima of the National Institute of Bioscience and Human-Technology in Tsukuba, Japan.

The precursor to the usability inspection method based on RAFIV was the Cockpit Cognitive Walkthrough (Polson and Smith, 1999), a method for structured design reviews that was derived from the Cognitive Walkthrough (Wharton, Rieman, Lewis, and Polson, 1994). The Cognitive Walkthrough is a widely used usability inspection method for evaluating ease of performing by exploration in office automation applications. The development of the Cockpit Cognitive Walkthrough was also supported by Cooperative Agreement, NCC 2-904 and carried out in collaboration with Nancy Smith (NASA Ames).

1.2 Outline and Purpose

This report is organized into seven sections. The next section describes the Goal-Situation-Behavior model and its applications. The following section summarizes the foundations of the RAFIV model and describes the model in some detail. The contents of both these sections are derived from previous reports referenced in footnotes. The fourth section integrates these two models into a complete design evaluation and training development methodology. The fifth section contains conclusions and possible future directions for research. References are in Section 6. Section 7 contains the titles and abstracts of the papers paper describing in more detail the results of this research program.

The main purpose of this report is to integrate the two models and provide the necessary context to characterize the accomplishments of this research program. The current efforts to validate these methods have involved showing that they can generate useful analyses of known operational and training problems with the current generation of avionics systems in modern commercial airliners (e.g.,

Eldredge, Mangold, and Dodd, 1992; Air Transport Association, 1997, 1998, 1999; BASI, 1998; FAA Human Factors Team, 1996).

2. The Situation-Goal-Behavior Model

The first method usability evaluation method developed during this research program was the Situation-Goal-Behavior (SGB) Model by Lance Sherry, Michael Feary and Everett Palmer (NASA Ames), and me. The SGB Model was derived from the Operational Procedure (OP) Model (Sherry, 1995). The OP model provides a formal description of the logic of an avionics system that is constructed so that it is comprehensible to avionics designers, pilots participating in system development, and training developers. In addition, the SGB model supports formal validation of the completeness and consistency of a design as well as automatic generation of the actual avionics software (Sherry and Feary, 2001).

2.1 Description of the Model¹

The Situation-Goal-Behavior (SGB) model is a rule based representation of an avionics system's functionality and behavior (Sherry, Feary, Polson, and Palmer, 2000c, Sherry and Feary, 2001). A situation is a particular pattern of possible inputs to the system including data from sensors and pilot actions recorded by various cockpit controls. The specific pattern of input values defined by the situation of a give rule is the trigger for the execution of that rule.

An action is represented by a pattern of values of outputs of the system. In case of a modern autopilot, the outputs are a control mode and values of the targets for the selected control mode as well as annunciations and other changes in cockpit displays and changes in the behavior of the aircraft.

A rule's goal is an intentional definition of its behavior. For example, a rule with the goal "climb and maintain a target altitude at a pilot selected rate of climb" would engage the vertical speed mode with specified target altitude and vertical

¹ The following section was paraphrased from Sherry, L & Feary, M. (2001) Improving the Aircraft Cockpit User-Interface: Using Rule-Based Expert System Models. *PC AI*, 15(6), 21-25.

speed. The collection of the goal-behavior descriptions for a system is a complete description of its functionality.

The SGB model is constructed from either an input-output analysis of the software, or derived from a flow-of-control analysis of the software algorithms. The model is an aggregated rule-based model with the same behavior as the actual software. The separation of the decision-making for determining the situations, from the algebraic computation/data manipulation of the actions, provides the mechanism to perform analyses described below. Also the assignment of situation-action pairs with goal labels provides a way to manage the complexity of the model. This definition represents the complete set of behaviors, or goals, of the system. (Sherry and Feary, 2001)

2.2 Applications of the SGB Model

Sherry and collaborators have demonstrated several different uses of the SGB model in the development of new avionics systems and for the analysis of existing systems in order to develop better training and reference materials.

They have performed SGB analyses on the pilot interface to a modern autopilot (Sherry, Feary, Polson, Palmer, 2000b, Sherry, Feary, Polson, Palmer, 2001c) and on the vertical guidance function of an FMS like those in the Airbus A320 and Boeing 777(Sherry, Feary, Polson, and Palmer, 2001b). These analyses examine the relationships between individual rules to uncover identical pilot actions that trigger different behaviors and different behaviors with missing or identical feedback to pilots.

They found the following kinds of flaws in these systems. Modal controls are where a given pilot action is context dependent and will invoke different behavior-goal combinations in different situations. Another problematic situation for pilots is where the normal action of a control is inhibited with out any indication why it was inhibited. Problems with feedback to pilots include the same annunciation for different goal-behavior pairs and different annunciations for the same behavior.

The other two applications of a SGB model exploit the fact that it is an alternative representation of a very complex software system. The behavior and the triggering situation for each possible goal-behavior combination are described in an executable formalism. The inputs triggering situations, behaviors, and goals are all described in terms meaningful to pilots.

Sherry and his collaborators (Feary, Sherry, Polson, and Palmer, 2000b; Sherry, Feary, Polson, and Palmer, 2001b) built a tutor for a modern autopilot that enables a pilot to see and explore the consequences of his actions during a simple loft. The tutoring system presented a detail representation of a mode control panel (MCP) and a primary flight display (PDF) augmented with additional information and other windows showing the details of the autopilot logic. Pilot actions generated the same changes in the MCP and PDF that would occur in an actual cockpit, e.g., changes in the flight mode (FMA) annunciator plus additional information, not available in real cockpits, that made transparent the underlying logic of the autopilot.

The other application was a reference tool. Sherry and collaborators (Sherry, Feary, Polson, Mumaw, and Palmer, 2001a) performed an SGB analysis of the vertical guidance function of a modern Flight Management System (FMS). They then incorporated it to a running simulation that included all relevant cockpit displays and controls (MCP, PDF, and MCDU) and a window showing details of the system logic. There were also controls that enable the user to manipulate the simulation's state and interact with the cockpit controls and displays. The function of the reference tool was to enable developers of reference and training materials to obtain a complete and correct understanding of the underlying vertical guidance logic and the relationships between situations, pilot actions, and changes in cockpit displays.

Sherry has also carried out an informal study of training and reference documentation developed by various airlines for both autopilots and vertical guidance functions of FMSs. In almost all instances, he found descriptions that seem to be derived from incomplete or incorrect understanding of the underlying system logic.

2.3 Summary

An SGB model of a system describes it as a collection of rules where each rule represents a situation and a behavior-goal pair that is triggered when a pattern of input values defining that situation occurs. The published applications of the model have all involved analyses of existing avionics subsystems, e.g., an autopilot and vertical guidance. In all cases, these analyses uncovered problems with pilot interfaces and poorly understood details of a system's logic.

Ongoing research using the SGB model involves applying it to the design of new systems. The objective of this research is to show that the SGB model is a

practical tool for use in actual avionics development environments and that the resulting systems will have better pilot interfaces and more transparent underlying logic. The ultimate goal is to design systems that reduce training costs and enhance operational efficiency and safety.

However, the SGB model is just part of a complete description of the interactions between pilots and the avionics. Another model is necessary that describes a pilot's knowledge and skills that enables her to use the autoflight system to accomplish tasks and to monitor performance of tasks by the automation. The RAFIV model describes the cognitive processes of a pilot using the autoflight systems. The model described in the next section focuses on the programming of the FMS using the Multifunction Control and Display Unit (MCDU).

3. The RAFIV Model

RAFIV is an approximate, five-stage model of pilot interactions with the autoflight system (Sherry, Polson, Feary, and Palmer, 2002a; Sherry, Polson, Fennell, and Feary, 2002b). The model describes the mental operations and physical actions of a pilot while performing a task (e.g., responding to an ATC clearance) using one or more subsystems (e.g. FMS or autopilot) of the autoflight system. RAFIV is an acronym for the five stages in the model, which are Reformulate, Access, Format, Insert, and Verify. Each stage is classified as either a recognition stage or a recall stage based on the interactions between the pilot and a system interface required to perform a stage. This classification plays a crucial role in usability analyses applications of RAFIV. The version of RAFIV described here focuses on pilot interactions with the Multifunction Control and Display Unit (MCDU) and other cockpit displays including the PDF, FMA, and Navigation Display (ND).

This section begins with short descriptions of the foundations of the RAFIV model and then describes the usability evaluation method based on the model and the details of each stage. Then briefly summarizes previous studies of operational and training problems with the FMS. Next the results of analyses based on the RAFIV model are presented, and the cognitive processes underlying each stage are discussed in more detail. The last two sections discuss the implications of the results of the RAFIV analyses for the development of more effective training programs for the FMS.

3.1 Foundations

3.1.1 Theory²

The theoretical foundations for RAFIV model are action-planning models of skilled performance (Kitajima & Polson, 1995) and learning by exploration (Kitajima & Polson, 1997). Both are models of display-based cognition that have evolved from previous models (Hutchins, Hollan, & Norman, 1986; Larkin, 1989). They are also consistent with both Zhang's (1997) and Hutchins' (1995) analyses of distributed cognition.

The Kitajima and Polson framework is based on Kintsch's (1998) construction-integration cognitive architecture originally developed to model skilled reading. The major claim of The Kitajima-Polson framework is that the action planning process is controlled by interactions between a pilot's representation of a task (i.e., goals) and labels on cockpit displays and controls.

In the case of the MCDU, this information includes key labels, the current page title, line labels, and line contents. The other critical component is pilots' knowledge of the interface conversions of the MCDU, the functions of page/mode keys, how to enter and edit information in the scratch pad, and that "<" or ">" at the beginning or end of a line-select key label on the CDU display is a soft page/mode key.

The Kitajima-Polson framework focuses on relationships between a pilot's goals, the actions available to the pilot, and the current state of the cockpit displays. Matching elements of a goal to labels on controls and displays enables the model to connect representations of goals, actions, and displays. Correct actions that are directly linked to goals and displays will be learned rapidly and performed correctly even in high workload situations. Because such connections are not retrieved from long-term memory, the model predicts that these actions will be error resistant.

The RAFIV model classifies a stage whose actions are generated by direct links between pilot goals and labels and other cues on controls and displays as a Recognition Stage. Thus, the Kitajima-Polson framework predicts and Franzke's (1995) data and numerous other studies support the claim that these recognition stages will be easily learned and performed correctly during line operations.

² This summary is derived in part from Polson (1999), "Final Report: A Comprehension Based Analysis of Autoflight System Interfaces", Section 2.

Correct actions whose links between goals and displays are retrieved from long-term memory will be harder to learn and error prone. The RAFIV model classifies a stage whose actions are generated by links retrieved from long-term memory between pilot goals and mismatching labels and other cues on controls and displays as a recall stage. Kitajima (1995) derives expressions for the probability of an error as a function of the number of links (fewer is better) and strengths (more is better) of the links between goals and correct actions. Franzke (1995) and Kitajima, Soto, and Polson (1998) showed that knowledge represented in these retrieved links can be difficult to learn and that these kinds of links are difficult to remember even after the learner has been explicitly tutored. Generalizing from the results of the Franzke (1995) and Kitajima, Soto, and Polson (1998) studies, we predict that actions necessary to complete a recall stage will be difficult to train and a source of errors during line operations because these links are easily forgotten.

3.1.2 Cockpit Cognitive Walkthrough³

A major goal of the previous research program was to generalize a theoretically based, usability evaluation methodology (Wharton, Lewis, Rieman, and Polson, 1994) developed for office automation to the modern, automated cockpit. This evaluation method was called the Cockpit Cognitive Walkthrough.

The Cockpit Cognitive Walkthrough is a design and evaluation method for prototype interfaces, cockpit procedures, and training materials for glass cockpit aircraft. It is a usability inspection method (Nielsen & Mack, 1994) that evaluates interactions between a cockpit procedure and an interface to an avionics system. The objective of the method is to show that the interface supports execution of the procedure under evaluation by providing feedback for correct pilot actions and error recovery, and by guiding the execution of a novel or an infrequently performed procedure. This focus on providing adequate support for exploration also improves other attributes of usability, including ease of learning and ease of use.

The Cockpit Cognitive Walkthrough is based on the cognitive walkthrough (Wharton, Lewis, Rieman, and Polson, 1994). The cognitive walkthrough evaluates the ability of a skilled user of an environment like the MAC OS to perform novel or occasionally performed tasks by exploration. The analysis is

³ This summary is derived from Polson (1999), "Final Report: A Comprehension Based Analysis of Autoflight System Interfaces", Section 4

very fine grained at the level of individual actions and cognitive operations. The evaluation criteria are derived from a theory of performing a task by exploration (Kitajima and Polson, 1995, 1997). Both versions of the Cognitive Walkthrough evaluate a system interface's support of perform by exploration, and both methods can be used by designers and developers who do not have formal training in human factors

3.2 Usability Evaluations Based On RAFIV

Sherry and his collaborators have adapted the Cognitive Walkthrough evaluation method to incorporate the recall and recognition terminology of RAFIV and combined it with the model to evaluate autoflight systems.

The Sherry et al. (2002a, b) analysis of the MCDU assumes that a pilot has knowledge of the interface conventions of the MCDU and the basic skills necessary to manipulate its controls. In addition, they assumed she/he is a licensed commercial airline pilot with all of the knowledge and skills necessary to fly an aircraft understand and correctly respond to clearances, and so on. What the RAFIV analysis of a task accesses is the additional knowledge that is required to carry out these tasks by programming the FMS using its MCDU interface, i.e. the linking knowledge described by Kitajima and Polson (1997).

RAFIV is an approximate model that supports the rapid usability evaluation of pilot tasks involving MCDU interface. This new method groups the sequence of mental operations and physical actions necessary to complete a task using the MCDU into five stages with each stage representing a major subtask that has to be accomplished by a pilot during execution of a task (e.g., responding to an ATC clearance). The analysis is at the level of subtasks (stages) rather than individual actions.

The focus of the analysis is on how cues provided by the system interface (e.g., function key labels, page titles, and line labels) support generation of actions necessary to successfully execute a stage. A stage is categorized as recognition stage if components of a pilot task description match labels on the system interface. For example, the label on the HOLD function key matches the ATC command in a pilot's description of the task of responding to a hold clearance. When there are no matches, the stage is classified as a recall stage. Usability analyses using methods based on the RAFIV model enable an avionics designer or developer to evaluate a large number of pilot tasks (over 100), given the time and other resource constraints imposed by the typical system development process.

The usability inspection method derived from the RAFIV model also incorporates two other innovations that are necessary to make it applicable to the analysis of the interactions between pilots and cockpit interfaces to the autoflight system. The first is the Reformulate stage. The processes underlying Reformulate were first described by Palmer, Hutchins, Ritter, and VanCleemput (1992). This stage models the pilot's transformation of her understanding of a task (e.g., responding to an ATC clearance) into a set of functions that can be performed by the automation. In most office automation application, there is a specialized function that will perform the task. Sherry et al. (2002,a,b) showed that this is not true for autoflight systems and this is why many tasks require the execution of complex reformulate stages.

The second is the Verify stage. The pilot is delegating to the autoflight system responsibility for execution of a task (e.g., climb to and maintain a specific altitude). The pilot must verify that the automation has been properly programmed to execute the task and then must monitor the automation's performance of the task. Thus, the Verify stage for cockpit automation is much different and more complex than just evaluating the system's feedback in response to the last command in an office automation application.

3.2.1 Recall and Recognition

The distinctions between recall and recognition stages are based on the principles underlying the design of applications with a graphical user interface that were proposed by the developers of the Xerox Star (Smith, Irby, Kimball, Verplank & Harlem, 1982), the precursor to the modern Microsoft Windows and Macintosh GUIs. Paraphrasing Smith et al., the pilot should be able to *recognize and select* the appropriate action from a set of alternatives presented by the interface rather than *recalling and typing*, the latter being more time-consuming and error-prone. Recognizing and selecting is a brief description of the primary cognitive process that supports performing by exploration.

Sherry et al. claim that training and operational problems with the FMS are caused by recall stages because they are difficult to acquire and remember as compared to recognition stages. They supported these claims about recall steps by citing studies of the learning and retention of new procedures for office automation applications (Kitajima, Soto, and Polson, 1998) and a FMS training study by McLennan, Irving, Polson, and Blackmon (Submitted).

3.2.2 Stages of the Model

The following summary of the RAFIV model has been paraphrased from the description in Sherry et al. (2002b). It briefly characterizes each stage of the model and describes attributes that would cause it to be classified as a recall stage.

Reformulate a task (e.g., responding to an ATC clearance) into a description of the function (or feature) of the automation that will perform the tasks, and the parameters of the function that must be entered into the MCDU. For example, an ATC clearance Direct-to a waypoint can be executed using the DIRECT TO feature of the MCDU LEGS page. The waypoint is the data that must be entered. The Reformulate stage is classified as a *recall stage* when the pilot must: (a) convert the mission task to a representation supported by a feature of the automation (e.g., non-standard clearance, or workaround), (b) compute data for entry to the automation (e.g., mental math), and (c) recall the existence of an automation feature to support the mission task (e.g., hidden feature).

Once a description on how to use the automation has been generated, the pilot must transfer the description to the automation via a sequence of actions that can be divided into three stages (Polson, Irving, and Irving, 1995).

Access the correct page to display the fields for data entry by selecting MCDU mode keys and/or navigating the hierarchy of pages on the MCDU. An Access stage is classified as a *recall stage* when the pilot must recall a memorized action sequence to get to the right page.

Format data for entry according to formats accepted by the MCDU pages. Most formatting takes place while typing entries into the scratchpad. The Format stage is classified as *recall stage* when the pilot must: (a) recall the format (e.g., the entry of a lateral route offset on the MCDU is <Side L or R><distance in nm.>), or (b) recall the content of the entry (e.g. ICAO indent for a waypoint in ATC instruction).

Insert data in the correct location using the Line Select keys on the MCDU to insert data typed into the scratchpad. The Insert stage is classified as a *recall stage* when the pilot must choose a Line Select key to insert the entry by recalling the correct location from memory.

The format and insert stages maybe repeated several times for multiple entries (e.g., responding to a complex hold clearance).

Verify & Monitor that the automation: (1) accepted the pilot entry, (2) is performing the intended task within the envelope of acceptable performance, and (3) the task is satisfying the mission goals (Fennell, 2002). This step involves visual scan and intensive scrutiny of the primary flight display (PFD), navigation display (ND), and MCDU. This stage is classified as *recall stage* when any of this information must be inferred from the displays.

3.3 How RAFIV Accounts for Results and Conclusions from other studies

Numerous studies starting with those by Curry (1985) and Wiener (1988) have documented operational and training problems with the modern autoflight system, in particular the Flight Management System (FMS) and its pilot interface, the MCDU (Eldredge, Mangold, and Dodd, 1992; Air Transport Association, 1997, 1998, 1999; BASI, 1998; FAA Human Factors Team, 1996). To account for these problems, various researchers have focused on different aspects of the autoflight system including its complexity (e.g., Billings, 1997; Sarter and Woods, 1995), lack of pilot understanding of its underlying logic (e.g., Sarter and Woods, 1997; Sherry and Polson, 1999), and various limitations of the MCDU interface (Woods, Johannesen, Cook, and Sarter, 1994). All of the preceding articles and reports as well as numerous other publications supporting these conclusions are based on data from accident and incident reports, experiments in simulators, and analyses of the underlying logic and pilot interfaces of the automation.

3.3.1 Results From An Analysis of the Boeing 777 MCDU

Sherry and collaborators (Sherry, et al, 2002a, b) have used the usability inspection method derived from the RAFIV model to analyzed 102 airborne pilot tasks (e.g., responding to ATC clearances) described in the Honeywell Pilot's Guide for the Boeing 777 FMS (Honeywell, 1995).

Each task was broken down into the five stages in the RAFIV model, and then each stage classified as a recognition or a recall stage (Sherry et al., 2002b). They also estimated the frequencies with which tasks occurred during normal line operations.

The follow is a description of the consequences of the large number of recall stages found during the RAFIV analysis of 102 tasks using the FMS in the Boeing 777. (Sherry et al. 2002b, pp. 1-2)

Investigations by researchers of modern flight-deck operations have identified the complexity of learning and using functions provided by the Flight Management Computer (FMC) and it's the Multi-function Control and Display Unit (MCDU). Pilots described the experience of learning to use this automation as "drinking from a fire hose" (BASI, 1999; p. 38), and only achieve skilled and efficient use of the system after 12 to 18 months of line experience (Polson, Irving, Irving, 1994). Several studies and surveys of pilots have consistently revealed that pilots have difficulty in using the features of the MCDU during line operations due to gaps in their knowledge (Mumaw, Sarter, and Wickens, 2001); Air Transport Association 1998; FAA 1996; Feary et al. 1998) and cite the need to more training (Air Transport Association, 1997, BASI 1999; p. 38).

Issues with learning and using the FMC/MCDU have been attributed to the lack of a detailed conceptual understanding of how traditional pilot tasks are performed by the FMS/MCDU (Sarter and Woods, 1992; Bobbitt, 2001). Other researchers have discussed the awkward layout of the keyboard (Sarter and Woods, 1994), the number of pages and features (Billings, 1997) the complexity of navigating through the hierarchy of pages, and the inefficiencies in inputting data (Abbott, 1997).

Whereas all of these phenomena contribute in varying degrees to the perceived complexity of the device, ultimately it is the number of memorized action sequences that must be learned and then recalled during high-tempo, safety critical line-operations that determine the ease of training and operation. This paper describes an analysis of a sample of 102 mission tasks that can be performed using functions supported by the MCDU as described in a standard Pilot's Users Guide (Honeywell, 1995).

76 of 102 tasks (75%) required memorization of at least one step to complete the task.

This directly contributes to the “drinking from a fire-hose effect” experienced by pilots during transition training (BASI, 1999; page 67)

46 of 102 tasks (45%) are performed infrequently and require memorized actions sequences.

This directly contributes to the perceived gaps in pilot knowledge during line-operations and the desire for more training to maintain proficiency (Air Transport Association, 1997). These results have implications on the way this automation is trained, and on the way future generations of the automation should be designed and certified.

3.3.2 Burdens Place on Memory

Current training programs for fight automation inadvertently place pilots in the position of having to rote memorize action sequences identified by Sherry et al (2002b). A typical CBT lesson starts with a description of the task and an example in a realistic flight context (e.g., an ATC clearance) followed by a list of actions required to perform the task. Reference materials like the Honeywell FMS Pilot’s Guides and airline Flight manuals simply list the actions required to execute a task with diagrams showing the state of the MCDU display at various stage in the procedure. In addition, the Reformulate stages are not described for any task. Versions of the Verify stage are included for some tasks trained in CBT lessons. Skilled pilot instructors do emphasize Verify stage techniques (Fennell, 2002).

Sherry et al. (2002b, quote above) argues that rote memorization is difficult and time consuming, and memorized procedures will be quickly forgotten unless reinforced by practice. In addition, the resulting skills are brittle. Pilots, especially those with limited glass cockpit experience, will not recognize novel variations of trained tasks or new tasks that can be performed using modification of already mastered procedures. They will also have trouble reconstructing a forgotten step in an infrequently performed procedure because they do not have the knowledge necessary to infer it.

3.3.3 What is Missing?

Comparing RAFIV analyses of tasks performed using the FMS with contents of training materials and reference documentation shows that many of the skills needed to operate the FMS are neither documented nor included in training. First, there are no systematic presentations of the Reformulate stage, the skills necessary to map a triggering event into a plan to use the FMS. Pilots are left to their own devices to discover systematic methods to perform the Reformulate stage. Second, no distinction is made with recognition and recall stages. Pilots can't focus their attention on the parts procedures that are going to be difficult to remember. Third, pilots not given any systematic support for memorizing actions required by recall stages.

3.3.4 RAFIV as a Tool for Training

The RAFIV model itself would enable pilots to organize the knowledge and skills necessary to perform all tasks that involved using the FMS by providing as descriptions of the major stages common of all tasks perform using the MCDU interface. Breaking up all of the information needed for a task into meaningful units would facilitate learning and later recall.

Training for the Reformulate stage should include the follow components. The first is a complete description of the functions of the FMS. Second, pilots should be taught specific reformulation skills for groups of related tasks to enable them to describe tasks in ways that better match the capabilities of the FMS (e.g., lateral flight plan modification.

Finally, pilots have to be taught skills that will enable them to deal with novel or infrequently performed tasks. A first step would be to identify and train general reformulation skills enabling pilots to identify a relation know task and generalize that procedure to dealing with the novel task. Second, the RAFIV model can be used as a memory aid to facilitate recall of infrequently learned procedures.

Sherry et al. concluded that recall stages in procedures are the final common path for operational and training problems caused by limitations of the MCDU interface and the underlying logic of the FMS that have been identified by other investigators. Recall stages are potential source of errors during line operations because they are easily forgotten for infrequently performed tasks. As a result, training for pilots becomes a highly time consuming and laborious process because of the large number of recall stages in the procedures they must learn for carrying out airborne tasks using the MCDU.

4. Combining RAFIV and SGB Models

The goal of this section is to integrate the two separate projects supported in part by this Cooperative Agreement. A SGB model is a representation of one of the subsystems of the autoflight system (e.g., autopilot, vertical guidance function of the FMS). The RAFIV model is an approximate description of the cognitive process of a pilot using a subsystem to perform a task. In the following, we describe the relationship of RAFIV to standard models of cognition and show how the RAVIF and SGB models are combined for complete usability analysis of a subsystem.

4.1 RAFIV and Models of Cyclic Interaction

Sherry, Feary, Polson, and Palmer (2001) show that the standard model of interaction with any form of automation (and much other human behavior) can be thought of as a continuous process of cyclic interaction between human operator and the system (Monk, 1999). This cyclic framework has a long history which includes Card, Moran, & Newell's (1983) recognize-act cycle in their Model Human Processor, Hutchins, Hollan, and Norman's (1986) analysis of direct manipulation, Kieras and Polson's (1985) formal analysis of user complexity, and the Kitajima and Polson framework (1995, 1997).

4.1.1 The Norman Approximate Model of Cyclic Interaction

Norman's (1988) approximate model of cyclic interaction that incorporates seven stages and is a summary of all of these models as well as many other formal models of human computer interaction (Byrne, 2003). In following, we will refer to the whole class of cyclic, human-computer interaction models as the cyclic interactions models include all of the models referenced in the preceding

Starting with a system's response to the last pilot action, Norman's (1988) model describes seven cognitive steps arranged in a loop: 1) perception of changes of system state, 2) comprehending the changes, 3) evaluating them in light of a pilot's goals, 4) modifying the goals if necessary, 5) formulating an intended result for the next action(s), 6) specifying the action(s), and 7) executing the action(s). The cycle time of this loop ranges from a fraction to several seconds and it can incorporate many mental operations and physical actions described by fine grain models of human-computer interaction (e.g., Kitajima and Polson, 1997). Most all cyclic interaction models implicitly or explicitly incorporate versions of these seven steps.

4.1.2 Stages

RAFIV is an approximate model of pilots' cognition that is at a different grain size than models of cyclic interaction. RAFIV is high-level description of pilots' cognition starting with the detection of a triggering event (e.g., an ATC clearance) to programming the FMC to monitoring the performance of the aircraft during the maneuver commanded by a pilot. RAFIV describes the sequence of subgoals (e.g., access the MCDU page associated with the task) defined by each stage during a pilot's interaction with the MCDU while performing a task. Each stage of the RAFIV model can involve one or more iterations through the steps described by a model of cyclic interaction.

The usability inspection method derived from the RAFIV model evaluates the properties of an interaction between a pilot and the MCDU as the pilot goes through the processes described in a model of cyclic interaction *for a specific stage*. If the correct actions can be generated by the processes described in the model from cues provided by the interface, then that stage is classified as a recognition stage. Otherwise, it is a recall stage. For example, Access is a recognition stage for the task of responding to a HOLD clearance using the FMS. There is a function key labeled HOLD on the MCDU interface. The match between the pilot's goal (e.g., respond to the HOLD clearance) and the label enables the processes described by the Norman loop to generate the correct action for the Access stage (e.g., press the HOLD key).

4.2 Three Loops

The purpose of this section is to characterize the relationships between the RAFIV model and more fine grain analyses of human-computer in terms of three nested control loops. The outer loop is called the Task Loop, and it is described as the sequences of stages, i.e., RAFIV. The middle loop is a model of cyclic interaction. The inner loop involves processing the immediate feedback from activation of keys, switches, and other controls.

RAFIV is the Task or outer loop. The cognitive activities range from detection and interpretation of the triggering event to programming the FMS to monitoring the behavior of the aircraft. The model is an approximate description of pilot-automation interaction whose time course can span from 1 to several minutes. In addition, the monitoring aspects of the Verify stage may be interleaved with the execution of other tasks (Polson and Javaux, 2001).

The middle loop has a time course from a few seconds to 30 seconds. The execution of a stage typically involves one iteration of a model of cyclic

interaction. Feedback is provided by the changes in the MCDU, ND, and other cockpit displays. A good example is execution of an Access stage where one or more pilot actions cause the MCDU page associated with a task to be displayed.

The inner loop involves the immediate tactile and visual feedback in response to operation of keys, switches, and other controls. This feedback is used to confirm that the last low-level physical action was successfully completed.

4.3 The Cockpit Verses Office Automation

The RAFIV model is an important contribution to our understanding of human-computer interaction in general and cockpit automation in particular. Models of user interaction with office automation have not explicitly described the Reformulate and Verify Stages. Models of users interacting with an application to perform a task (e.g., Kitajima and Polson, 1995,1997) have focused in describing these interactions at the level of a model of cyclic interaction.

The Cognitive Walkthrough or any other usability evaluation method for office automation applications does not incorporate evaluations of analogs of either the Reformulate or Verify Stages. An office application requiring complex Reformulation stages to perform task would be rejected on the grounds that it did not have functions that directly support important user tasks.

The Verify stage is not included in office automation models because most tasks are performed incrementally (e.g., small step by step changes to a manuscript, spreadsheet, or drawing). The user does not delegate the performance of typical tasks to the system. A model of cyclic interaction is a model of the incremental performance of a complex task.

4.4 Role of The Situation-Goal-Behavior (SGB) Model in Usability Analyses

Kieras and Polson (1985) argued that a complete formal analysis of the interactions between operator and system necessary to complete a task require two models, one of the cognitive processes of the operator (e.g., RAFIV) and a second that model the behavior of the system in responding to the operators actions. The SGB model is a high level description of the system's behavior.

An SGB model represents an avionics system's functionality and behavior as a collection of situation-behavior rules with a goal based description of the function of each rule. A rule is triggered by a pattern of input that match its situation pattern of inputs including inputs from sensors and pilot actions

recorded by various cockpit controls. The behavior describes the outputs of a system including changes in cockpit displays and behavior of an aircraft.

An SGB model provides critical inputs to a RAFIV analysis. The SGB model describes the collection of goals that can be accomplished by a system's behaviors. This model also represents the controls that generate inputs to the system and the displays it generates in response to pilot inputs and inputs from its sensors.

In the paragraphs that follow, we will show how a SGB model of system supports a RAFIV analysis of its usability.

4.4.1 Reformulate

Tasks performed by pilots in the cockpit are initiated in response to events. Examples are an ATC clearance, a checklist item, an SOP retrieved from memory, an alert or warning, or an abnormal condition detected by pilot (Polson and Smith, 1999). A pilot must understand the initiating event in order to formulate a description of the task to be performed in response to it. Correct interpretation of an alert or warning or of a complex or novel clearance is obviously a necessary initial stage in successful use of the automation

The Reformulate stage involves generating a plan to use the automation to perform the task defined by the initiating event. A RAFIV analysis of the Reformulate stage focuses on complexity of the transformations required to map the pilot's understanding of the triggering into one or more functions described by the SGB model. An SGB model describes the various functions of the automation that are available to accomplish the task defined by the triggering event.

In short, Reformulate is a mapping from the pilot understanding of the task to the functions of the automation described by the SGB model.

The complexity of this transformation determines the difficulty of learning how to do the task with the automation and how well the interface supports the pilot's programming of the automation to carry out the task. If the task's representation must be extensively elaborated transforming it into representations of the functions that can be performed by the FMS, the Reformulate stage is classified as a recall stage. An example would be a complex lateral flightplan modification requiring editing of both the RTE and LEGS pages.

Sherry et al. (2002a) showed that many of the operational and training problems identified by researchers like the limitations of VNAV (BASI, 1999) can be

explained by the complexity of the Reformulate stage for vertical navigation tasks like satisfying speed and altitude constraints at a fix. There is no single SGB rule in the FMS that directly performs this task. Thus, pilots have to reformulate such tasks into a complex sequence of FMS functions that only indirectly support compliance with such clearances. Responding to a HOLD clearance on the other hand can be accomplished by invoking a single rule that directly supports performance of the task.

In summary, a SGB analysis is the critical initial step in analyzing the Reformulate stages for a collection of pilot tasks performed by a system. A SGB model enables the usability evaluation team to understand how the limitations of a system's underlying functionality are impacting its usability. *A one task to many automation functions mapping is almost certain to lead to serious usability problems.*

4.4.2 Access, Format, and Insert Stages

Access, Format, and Insert all involve a least one iteration each through the Norman loop interacting with the controls provided by the pilot interface and interpreting the feedback generated by the system in response to each pilot action. The top level structure of a pilot's cognitive processes is described by the subgoals that characterize each of these three stages. These subgoals represent a pilot's understanding of the MCDU interface conventions and the organization of any task performed using the FMS.

The relationships between the pilot actions, the SGB rule(s) invoked by those actions, the resulting behavior of the aircraft, and the feedback presented to the pilot are all described by an SGB model. Making distinctions between recall and recognition stages requires a description of a pilot interface at the level provided by the SGB model. The labels on and effects of controls and the content of feedback provide by cockpit displays all determine the classification of a stage. In addition, the kinds of design errors uncovered using an SGB model (e.g., multiple functions invoked a single control and ambiguous or incomplete feedback) are all serious interface design errors in their own right. However, they also lead to recall stages in RAFVIV models of tasks that require their use.

4.4.3 Verify

The analysis of the Verify stage also depends on the description of the cockpit displays provided by a SGB model. For example, Sherry et al (2001) and Sherry and Polson (1999) show that the feedback provided by the vertical guidance

function of the FMS is ambiguous. The same vertical mode annunciation is displayed for rules with very different goals and behavior.

5. Summary and Future Research

Kieras and Polson (1985) concluded that a complete formal analysis of the properties of users interacting with an application requires both models of users and of the system users are interacting with. Sherry and his collaborators have developed a model of pilot interacting with the autoflight system (RAFIV) and models of various subsystems interfaces and logic (SGB). Their analyses showed various usability problems identified by other investigators all have a final common path, recall stages. Further, they showed that a large majority of tasks involved use of the MCDU have one or more recall stages explaining the known training and operational problems that have been identified by different investigators (e.g., Billings, 1997).

On going and future research on SGB and RAFIV models is taking four related tracks. The first is a theoretical track developing more detailed models of the cognitive processes underlying the five RAFIV stages. The second is the collection of new data to test the learning and perform predictions of the RAFIV model. The third is the development of training programs based on RAFIV and SGB analyses of cockpit procedures for aircraft like the Boeing 777 and the Airbus A320. The fourth track is the application of SGB and RAFIV analysis to new cockpit designs.

5.1 The Theory Track

In collaboration with Dr. Michael Matessa, Polson is developing detailed computer simulations of pilots responding to ATC clearances based in the RAFIV analysis of pilot interaction with the MCDU. The models will be developed using the latest version of the ACT-R cognitive architecture, ACT-R/PM (Anderson, Bothell, Byrne, and Lebiere, Under review; Anderson and Lebiere, 1998).

The ACT-R/PM architecture provides the theoretical tools necessary to answer question about details of the Sherry et al. recall and recognition stages. Successful development of ACT-R/PM simulation models of responding to clearances like hold and intercept a leg to will enable us to develop a much better understanding of recall and recognition stages. These simulation models will provide detailed descriptions of the cognitive mechanisms involved in both recognition and recalls stages for tasks that involve the programming of the FMS through the

MCDU. These detailed descriptions will support further development of the RAFIV usability inspection method. For example, it will be possible to specify how features of the MCDU interface can be exploited eliminate recall stages. The quantitative submodels in the ACT-R/PM architecture describing the acquisition and forgetting of facts and skills will enable us to quantify the costs of recall stages.

5.2 The Empirical Track

Definitive tests of the RAFIV predictions require three kinds of data. The first is information about the frequencies of tasks performed using the FMS during normal flight operations (e.g., intercept a radial to a VOR). The second is detailed observations tracing time courses of acquisition of a wide range of tasks. RAFIV predicts that the time required to train a task to proficiency will be an increasing function of the number and complexity of the recall stages in a task. The third kind of data is information about problems performing tasks using the FMS during normal line operations.

In collaboration with F/O Karl Fennell, a pilot instructor and the United Airlines Training Center, studies are being designed to test the RAVIF model's predictions using questionnaire data from several different groups of pilots employed by a major airline including line pilots, Line Check Airmen (LCAs) and Standards Captains. Each questionnaire will have two sections. The first will ask for information about relevant flying experience including current position, 777 flight hours, previous experience with other glass cockpit aircraft, etc. The second section will contain between 30 and 40 task descriptions followed by two to four rating scales. Pilots will be asked to estimate task frequency during line operations, rate training difficulty, and rate the likelihood of operational problems.

The advantages of using survey data to evaluate RAFIV predictions are that it is possible to quickly obtain frequency and difficulty estimates on a large number of tasks in a short period of time. In addition, all of the ratings we are asking for are based on information related to pilots' domain of expertise, flying a commercial airliner. When ever possible, attempts will be made to cross check the rating scale data with other measures.

5.3 Training Track

In Sections 3.4.3 and 3.4.4, it was shown that many of the skills needed to operate the FMS are neither documented nor included in training. First, there are no

systematic presentations of the Reformulate stage. Second, no distinction is made between recognition and recall stages. Third, pilots are not given any systematic support for memorizing actions required by recall stages.

Use of both the RAFIV and SBG models to design training materials could make dramatic improvements in the efficiency, effectiveness, and breadth of automation training. Current training programs force pilots to rote memorize many of the action necessary to operate the FMS. Such training is both time consuming and inefficient because rote memorized material is rapidly forgotten.

Systematic use of SGB models could dramatically improve pilots understanding of the logic of various systems and the details and pitfalls (e.g., modal controls). Such training would also solve a major operational problem, pilots' limited understanding of the autoflight system's logic (e.g., Sarter and Woods, 1995, 2000).

5.4 Design Track

Lance Sherry and Michael Feary are collaborating with several individuals at Boeing on the design of new autoflight system using both SGB and RAFIV models.

6. References

- Abbott, T (1997) *A Comparison of Two Control Display Unit Concepts on Flight Management System Training*, NASA Technical Memorandum 4744.
- Air Transport Association. (1997). *Towards an Operational Philosophy and Model Training Program for FMC-Generation Aircraft* (First report by the Human Factors Committee Automation Subcommittee). Request copy of this report by e-mail to Tom Chidester (tchidester@mail.arc.nasa.gov).
- Air Transport Association. (1998). *Potential Knowledge, Policy or Training Gaps Regarding Operation of FMS-Generation Aircraft* (Second report by the Human Factors Committee Automation Subcommittee). Request copy of this report by e-mail to Tom Chidester (tchidester@mail.arc.nasa.gov).
- Air Transport Association. (1999). *Performance of Standard Navigation Tasks by FMS-Generation Aircraft* (Third report by the Human Factors Committee Automation Subcommittee). Request copy of this report by e-mail to Tom Chidester (tchidester@mail.arc.nasa.gov).

- Anderson, J. R., and Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Anderson, J.R., Bothell, D.B, Byrne, M.D., Lebiere, C. (Under review). Integrated Theory of the Mind. *Psychological Review*.
- BASI (1999) Advanced Technology Safety Survey Report. Flight Safety Digest Special Issue. Flight Safety Foundation, June-Aug 1999, pages 137-216. Available at www.basi.gov.au/sprog/advtek.htm (9/00)
- Bellotti, V.M.E. (1990) A framework for assessing applicability of HCI techniques. *Proceeding of Interact90, 3rd IFIP Conference on Human- Computer Interaction*, Cambridge, England, August, 1990.
- Billings, C. E. (1997). *Aviation automation: The search for a human-centered approach*. Mahwah, NJ: Lawrence Erlbaum.
- Bobbitt, R. (2001) personal communication.
- Byrne, M. D. (2003). Cognitive architecture. To appear in J. Jacko & A. Sears (Eds.), *Human-Computer Interaction Handbook*. Mahwah, NJ: Erlbaum.
- Card, S.K., T.P. Moran, & A. Newell (1983). *The psychology of human-computer interaction*. Lawrence Erlbaum: Hillsdale, N.J.
- Curry, R.E. (1985). The introduction of new cockpit technology: A human factors study. NASA TM 86659. Moffett Field, CA: NASA Ames Research Center
- Eldredge, D., Mangold, S., & Dodd, R. S. (1992). *A review and discussion of flight management system incidents reported to the Aviation Safety Reporting System* (Final report no. DOT/FAA/RD-92/2). Washington, DC: Battelle/U.S. Department of Transportation.
- FAA Human Factors Team (1996). *Federal Aviation Administration human factors report on the interfaces between flight crews and modern flight deck systems..* Washington, DC: Federal Aviation Administration.
- Feary, M., Polson, P. G., Alkin, M., McCrobie, D., Sherry, L., & Palmer, E. (1998) Aiding Vertical Guidance Understanding. In G. Boy, C. Graeber, & J-M Robert (Eds.) *Proceedings of HCI-Aero-98, International Conference on Human Computer Interaction in Aeronautics* (pp. 165 - 170) Montreal: Ecole Polytechnique de Montreal.

- Feary, M., Sherry, L., Palmer, E., & Polson, P. G. (2000) Evaluation of a formal methodology for developing aircraft vertical flight guidance training material. *Proceedings of HCI-Aero-2000, International Conference on Human Computer Interaction in Aeronautics*. (pages 123 - 129) Toulouse, France: Cepadues-Editions
- Federal Aviation Administration (FAA) Human Factors Team. 1996. Report on the Interfaces between Flightcrews and Modern Flight Deck Systems (June 18, 1996). Washington: U.S. Department of Transportation, Federal Aviation Administration.
- Fennell, K. (2002) Personal communication.
- Franzke, M. (1995) Turning research into practice: characteristics of display-based interaction. *Proceedings of the CHI '95 Conference on Human Factors in Computer Systems*. New York, NY:ACM.
- Honeywell (1995) Boeing 777 Flight management System Pilot's Guide. Honeywell Publication # C28-3641-22-00. Honeywell International. Air Transport Systems, Product Support Distribution. PO Box 21111, Phoenix, Arizona, 85036-1111
- Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, 19, 265-288.
- Hutchins, E. L., Hollan, J. D., & Norman, D. A. (1986). Direct manipulation interfaces. In Norman, D. A. & Draper, S. W., Eds. *User Centered System Design*. 87-124. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Kieras, D.E., & Polson, P.G. (1985). An approach to the formal analysis of user complexity. *International Journal of Man-Machine Studies*, 22, 365-394
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Kitajima, M., & Polson, P. G. (1995) A comprehension-based model of correct performance and errors in skilled, display-based human-computer interaction. *International Journal of Human-Computer Studies*, 43, 65-99.
- Kitajima, M. (1995). Action selection error probabilities in the construction-integration theory-based model of display-based human-computer interaction. *ICS-TR*, 95-01. Institute of Cognitive Science, University of Colorado.

- Kitajima, M., & Polson, P. G. (1997). A comprehension-based model of exploration. *Human-Computer Interaction*. 12, 345 – 389.
- Kitajima, M., Soto, R., & Polson, P. G. (1998) LICAI+: A comprehension-based model of the recall of action sequences. In F. Ritter & R.M. Young (Eds.) *Proceedings of Second European Conference on Cognitive Modeling*. pp. 82-89. Nottingham University Press: Nottingham.
- Larkin, J.H. (1989) Display-based problem solving. In Klahr, D. & Kotovsky, K., Eds. *Complex Information Processing: The Impact of Herbert A. Simon*. 319-342. Hillsdale, New Jersey: Lawrence Erlbaum Assoc.
- Mangold, S. J., & Eldredge, D. (1995). Flight management systems information requirements. In R. S. Jensen (Ed.), *Proceedings of the Eighth International Aviation Psychology Symposium Conference*. Columbus: Ohio State University Press.
- McLennan, S., Irving, J. E., Polson, P. G., and Blackmon, M. H. (Submitted) *International Journal of Aviation Psychology*. Experimental Evidence Favoring Adoption of Cognitive Tutors for Flightcrew-Automation Training.
- Mumaw, R.J., Sarter, N.B., and Wickens, C.D. (2001). Analysis of pilots' monitoring and performance on an automated flight deck. *Proceedings of the 11th International Symposium on Aviation Psychology*. Columbus, OH: The Ohio State University.
- Nielsen, J. & Mack, R. (1994). *Usability Inspection Methods*. New York: John Wiley & Sons.
- Norman, D.A. (1988). *The psychology of everyday things*. New York: Basic Books.
- Palmer, E., E. Hutchins, R.D. Ritter, I. VanCleemput (1992) *Altitude Deviations: Breakdowns of an Error Tolerant System*. NASA/TM DOT/FAA/RD-92/7
- Polson, P. G., Irving, S., & Irving, J. E. (1994). Final report: *Applications of formal methods of human computer interaction to training and the use of the control and display unit*. Washington, DC: System Technology Division, ARD 200, Department of Transportation, FAA.
- Polson, P., & Smith, N. (1999) The cockpit cognitive walkthrough. *Proceedings of the Tenth Symposium on Aviation Psychology*. (pp. 427-432) Columbus OH: Ohio State University.

- Polson, P.G., Irving, S., and Irving, J.E. (1994) Final Report: Applications of Formal Models of Human-Computer Interaction to Training and Use of the Control and Display Unit. System Technology Division, ARD 200, Dept. of Transportation, Federal Aviation Administration. Also published as *ICS Technical Report*. 94-08. Boulder, CO: Institute of Cognitive Science, University of Colorado.
- Rieman, J., Young, R.M., Howes, A. (1996). A dual-space model of iteratively deepening exploratory learning. *International Journal of Human-Computer Studies*, 44, 743-775.
- Sarter, N. B., & Woods, D. D. (1992). Pilot interaction with cockpit automation I: Operational experiences with the flight management system. *The International Journal of Aviation Psychology*, 2(4), 303-321.
- Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode errors and awareness in supervisory control. *Human Factors*, 37, 5-20.
- Sarter, N.B. and Woods, D.D. (1997). "Team play with A Powerful and Independent Agent": A Corpus of Operational Experiences and Automation Surprises on the Airbus A-320. *Human Factors*, 39(4), 553-569.
- Sarter, N.B. and Woods, D.D. (2000). Teamplay with a Powerful and Independent Agent: A Full-Mission Simulation Study. *Human Factors*, 42(3), 390-402.
- Sherry, L. (1995). A formalism for the specification of operationally embedded reactive systems. In *Proceedings International Council on System Engineering*. St. Louis, MO.
- Sherry, L., & Polson, P. G. (1999). Shared Models Of FMS Vertical Guidance. *International Journal of Aviation Psychology*, 9, 139 - 153.
- Sherry, L. P. Polson, M. Feary, & E. Palmer (2002a) When Does the MCDU Interface Work Well? Lessons Learned for the Design of New Flightdeck User-Interfaces. Honeywell Publication C69-5370-0021.
- Sherry, L., P. Polson, K. Fennell, & M. Feary (2002b) Drinking from the Fire Hose: The over-reliance on memorized action sequences for MCDU operation. Honeywell Publication # C69-5370-022.
- Sherry, L & Feary, M. (2001) Improving the Aircraft Cockpit User-Interface: Using Rule-Based Expert System Models. *PC AI*, 15(6), 21-25.

- Sherry, L., Feary, M., Polson, P., Mumaw, R., and Palmer, E. (2001a). A cognitive engineering analysis of the Flight Management System (FMS) Vertical Navigation (VNAV) Function. *International Journal of Human Factors and Aerospace Safety*, 1, pp. 223-245.
- Sherry, L., Feary, M., Polson, P. G., & Palmer, E. (2001b) What's it doing now? Taking the covers off autopilot behavior. *Proceedings of the Eleventh Symposium on Aviation Psychology*. Columbus OH: Ohio State University. Also published as Honeywell Publications C69-5370-13
- Sherry, L., M. Feary, P. Polson, & E. Palmer. (2000a) Autopilot tutor: Building and maintaining autopilot skills. *Proceedings of HCI-Aero-2000, International Conference on Human Computer Interaction in Aeronautics*. (pages xx - yy) Toulouse, France: Cepadues-Editions. Also published as Honeywell Publications C69-5370-15.
- Sherry, L., M. Feary, P. Polson, & E. Palmer (2000b). Why won't it take that? MCP control device invokes more than one autopilot behavior. NASA-Ames Research Center, 2000. Also published as Honeywell Publications C69-5370-14
- Sherry, L., M. Feary, P. Polson, & E. Palmer (2000c). Formal Method for Identifying Two Types of Automation Surprises. NASA Ames Research Center. Honeywell Publications C69-5370-16.
- Smith, D. C. Irby, R. Kimball, B. Verplank & E. Harlem (1982). Designing the Star User Interface. *Byte*, vol. 7, no. 4, pp. 242-282.
- Smith, N., Moses, J., Polson, P., Romahn, S., & Palmer, E. (1999) A Survey of Flight Crew Experience with FANS-1 ATC Datalink. Presented at the 10th International Symposium on Aviation Psychology. Columbus, OH.
- Wharton, C., Rieman, J., Lewis, C., & Polson, P. (1994) The Cognitive Walkthrough Method: A Practitioner's Guide. In J. Nielsen and R. Mack (Eds.). *Usability Inspection Methods*. New York: John Wiley & Sons.
- Wiener, E. (1988). Cockpit automation. In E. L. Wiener & D. C. Nagel (Eds.), *Human factors in aviation* (pp. 433-461). San Diego: Academic Press.
- Woods, D.D., Johannesen, L., Cook, R.I., & Sarter, N.B. (1994). Behind Human Error: Cognitive Systems, Computers, and Hindsight. Crew Systems Ergonomic Information and Analysis Center (CSERIAC), Dayton, OH (State of the Art Report).

Zhang, J. (1997). The nature of external representations in problem solving.
Cognitive Science, 21(2), 179-217.

7. Publications, Reports, and Submitted Papers with Abstracts

Blackmon, M.B., Polson, P.G., McLennan, S.I., & Irving, J.E. (To be submitted)
Improving the effectiveness and efficiency of training for flight automation.
International Journal of Aviation Psychology.

McLennan, S.I., Irving, J.E., Polson, P.G., & Blackmon, M.B. (To be submitted)
Experimental Evidence Favoring Adoption of Cognitive Tutors for Flightcrew-
Automation Training. *International Journal of Aviation Psychology*.

Abstract: We created a low-cost, computer-based cognitive tutor following Anderson's eight guidelines for designing cognitive tutors derived from modern theories of cognitive skill acquisition (Anderson & Schunn, 2000). Then we tested pilots on all CDU tasks taught during training and mandated for the FAA checkride, using a realistic flight scenario in a full-motion simulator. Overall performance of experimentally trained pilots was equivalent to traditionally trained pilots who spent approximately six times longer in training on fixed-based simulators. The results provide a proof-of-concept for the feasibility of enlarging the flightcrew-automation training curriculum without increasing the length of current/traditional training and using inexpensive training equipment.

McLennan, S.I., Irving, J.E., Polson, P.G., & Blackmon, M.B. (Under review)
Improving Training on the Glass-Cockpit CDU Interface. NASA Ames Technical Report.

Abstract: We designed a 5-hour CDU training program based on the ACT-R theory of skill acquisition and associated principles for designing intelligent tutors. After training, experimentally trained pilots successfully completed all FAA-mandated CDU tasks in a full-motion simulator test comparable to the FAA checkride. Experimentally trained pilots' performance approximated that of traditionally trained pilots, who had spent 10-50 hours training on sophisticated simulators. Our training design can be applied to teaching flightcrews the full range of CDU tasks, and its time- and cost-efficiency demonstrates the feasibility of teaching substantially more CDU tasks and topics within current airline budgets for CDU training.

Blackmon, M.B. & Polson, P.G. (2002) Key Advances for Improving Aviation Training Design. S. Chatty, J. Hansman, & G. Boy (Eds.) *Proceedings of HCI-Aero-2002, International Conference on Human Computer Interaction in Aeronautics*.(Pp. 24-29) Menlo Park: AAI Press.

Abstract: Combining two recent technologies can markedly improve the performance outcomes and cost-effectiveness of aviation training. The first is a well-tested design methodology for developing cognitive tutors (Anderson et al. 1995, Anderson and Schunn 2000) based on modern theories of skill acquisition. The second is the advent of high-fidelity PC-based part-task simulators on which pilots can “learn by doing” and “progress to real-world performance,” two essential guidelines for designing cognitive tutors. An experimental flightcrew automation training program (McLennan et al. submitted) produced results consistent with non-aviation training results using Anderson’s cognitive tutors, implying that pilots trained on cognitive tutors can attain the same or higher level of competence in approximately one-third the training time for traditionally trained pilots.

Sherry, L. P. Polson, M. Feary, & E. Palmer (2002) When Does the MCDU Interface Work Well? Lessons Learned for the Design of New Flightdeck User-Interfaces. S. Chatty, J. Hansman, & G. Boy (Eds.) *Proceedings of HCI-Aero-2002, International Conference on Human Computer Interaction in Aeronautics*. (Pp. 180-185) Menlo Park: AAI Press. Also published as Honeywell Publication C69-5370-0021.

Abstract: The Multi-function Control and Display Unit (MCDU) has been identified as a source of issues pilots have transitioning to glass cockpits. Several aircraft manufacturers and avionics vendors have committed to replace the MCDU with graphical user-interfaces in the next generation of commercial aircraft.

A cognitive task analysis of pilot-MCDU interaction, described in this paper, has identified that pilot failure to complete mission tasks using the MCDU is not a sole consequence of the physical dimensions or layout of the device. Instead, the MCDU interface works adequately when a given pilot task: (1) is supported directly by a function provided by the automation, and (2) the access of MCDU pages, and format and entry of data, are prompted by labels and other visual cues (and not by memorized actions sequences). Pilot tasks not supported directly by automation, and/or pilots tasks that rely on memorized action sequences are difficult to learn and likely not to be used effectively in the field.

Sherry, L., P. Polson, K. Fennell, and M. Feary (2002) Drinking from the Fire Hose: The over-reliance on memorized action sequences for MCDU operation. Honeywell Publication # C69-5370-022.

Abstract: The Flight Management Computer (FMC) and its Multi-function Control and Display Unit (MCDU) have been identified by researchers and airlines as difficult to learn and use. This directly impacts the amount of functionality that is trained, the length of the training footprint, and the cost of training.

This paper examined the degree of reliance on memorized action sequences required to perform a sample of 102 mission tasks using features of the B777 MCDU. The analysis identified an over-reliance on memorized action sequences that must be learned during training and then recalled during line operations. This over-reliance directly explains the difficulties in learning and using the automation. Implications for training of these systems, and the design of new user-interfaces are discussed.

Sherry, L., Polson, P. and Feary, M. (2002) Designing User-Interfaces for the Cockpit: Five Common Design Errors, and How to Avoid Them. Paper to be presented at the 2002 SAE World Aviation Congress, Phoenix, AZ (November 5 – 7).

Abstract: The efficiency and robustness of pilot-automation interaction is a function of the volume of memorized action sequences required to use the automation to perform mission tasks. This paper describes a model of pilot cognition for the evaluation of the cognitive usability of cockpit automation. Five common cockpit automation design errors are discussed with examples.

Feary, M., Sherry, L., Polson, P., and Palmer, E.(2001). *A Formal Method for Integrated System Design*. 8th European Conference on Cognitive Science Approaches to Process Control (CSAPC '01), Munich, Germany (September 24-26).

Polson, P.G. & Javaux, D. (2001) A model-based analysis of why pilots do not always look at the FMA. *Proceedings of the Eleventh Symposium on Aviation Psychology*. Columbus OH: Ohio State University.

Abstract: Anecdotal reports and some studies show that pilots do not monitor the flight mode annunciator (FMA) when making routine mode changes on the mode control panel (MCP). These monitoring lapses are not intentional violations of SOPs. This paper describes the skill acquisition processes responsible for these changes in monitoring performance. Learning mechanisms modify the initially trained operating procedures that include careful monitoring

of the FMA. These modifications reduce the likelihood that pilots will fixate the FMA after making routine mode changes on the MCP.

Sherry, L., Feary, M., Polson, P., Mumaw, R., and Palmer, E. (2001). A cognitive engineering analysis of the Flight Management System (FMS) Vertical Navigation (VNAV) Function. *International Journal of Human Factors and Aerospace Safety*, 1, pp. xx-yy

Abstract: A cognitive engineering analysis of the Flight Management System (FMS) Vertical Navigation (VNAV) function has identified overloading of the VNAV button and overloading of the Flight Mode Annunciation (FMA) used by the VNAV function. These two types of overloading, resulting in modal input devices and ambiguous feedback, are well known sources of operator confusion, and explain, in part, the operational issues experienced by airline pilots using VNAV in descent and approach. A proposal to modify the existing VNAV design to eliminate the overloading is discussed. The proposed design improves pilot's situational awareness of the VNAV function, and potentially reduces the cost of software development and improves safety.

Sherry, L., Feary, M., Polson, P., Mumaw, R., and Palmer, E. (2001). *Cognitive Engineering Analysis of the Vertical Navigation (VNAV) Function*. NASA/TM-2001-210915. Ames Research Center, Moffett Field, CA, USA.

Smith, N., P. Polson, J.A. Brown, J. Moses, and E.A. Palmer (2001), An Assessment of Flight Crew Experiences with FANS-1 Controller-Pilot Data Link Communication in the South Pacific The Fourth International Air Traffic Management R&D Seminar ATM-2001, Santa Fe, NM.

Abstract: This paper presents "lessons learned" from a three-part human factors evaluation of the Future Air Navigation System's (FANS) controller-pilot data link communication (CPDLC) system. Three airlines have been using the CPDLC component of FANS since 1995, when air traffic facilities along South Pacific routes began providing data link communications services for FANS-equipped aircraft. This paper presents results from a fleet survey distributed to Boeing 747-400 flight crews with FANS CPDLC experience who fly for these three airlines. Two related activities—a task analysis-based usability evaluation of the FANS data link interface on the Boeing 747-400, and the collection and analysis of data link related reports submitted to NASA's Aviation Safety Reporting System—also contribute to the operational "lessons learned" presented in this paper.

Sherry, L., Feary, M., Polson, P. G., & Palmer, E. (2001) What's it doing now ? Taking the covers off autopilot behavior.. *Proceedings of the Eleventh Symposium on Aviation Psychology*. Columbus OH: Ohio State University. Also published as Honeywell Publications C69-5370-13

Abstract: Automation surprises in a modern "glass cockpit" can be attributed to the absence of a shared understanding of the intentions of the automation by the pilot/automation system. This paper demonstrates, using a formal modeling technique, how cockpit Flight Mode Annunciation (FMA) and Primary Flight Displays (PFD) fail to distinguish between distinct autopilot control behaviors. When the "covers are taken off" the autopilot behavior, it is observed that a single FMA configuration represents more than one autopilot behavior. This paper also describes how the contents of the formal model can be used to develop training and to set certification criteria.

Feary, M., Sherry, L., Polson, P., and Palmer, E.(2000). Formal Method for Developing Part-Task Training for a Modern Autopilot. Paper presented at the 3rd International Conference on Engineering Psychology and Cognitive Ergonomics, Edinburgh, Scotland (October 25-27).

Abstract: "Automation surprises" occur when operators of sophisticated automation, such as pilots of aircraft, hold a mental model of the behavior of the automation that does not reflect the actual behavior of the automation. This leads to increased workload, and reduced efficiency and safety. This paper describes a formal method for analysis of automation and it's user-interface for two well known characteristics that lead to automation surprises: (1) an automation user input device that, when selected, results in different automation behaviors depending on the situation, and (2) automation displays that do not provide unique annunciation for all automation behaviors. This method is unique in that it is based on analysis of the goals and behavior of the actual automation software. This provides a meaningful basis to perform user-oriented task analyses. A case study is also provided. Keywords: Automation surprise, formal methods, human factors, certification criteria.

Sherry, L., M. Feary, P. Polson, & E. Palmer. (2000) Autopilot tutor: Building and maintaining autopilot skills. *Proceedings of HCI-Aero-2000, International Conference*

on *Human Computer Interaction in Aeronautics*. (pages xx - yy) Toulouse, France: Cepadues-Editions. Also published as Honeywell Publications C69-5370-15.

Abstract: This paper describes a web-based tutor used to build and maintain pilot skills in operating a modern autopilot. The tutor, based on a goal-based model derived from the actual autopilot code, explicitly defines: (1) knowledge to recognize all unique autopilot behaviors from information on the flight mode annunciation (FMA) and other primary flight display (PFD) cues, (2) knowledge to convert pilot goals into pilot actions on the mode control panel (MCP). The tutor builds and maintains pilot skills by requiring the pilot to "solve problems" by executing Air Traffic Control instructions. The tutor provides immediate feedback to reinforce correct pilot behavior and rectify incorrect pilot behavior.

Feary, M., Sherry, L., Palmer, E., & Polson, P. G. (2000) Evaluation of a formal methodology for developing aircraft vertical flight guidance training material. *Proceedings of HCI-Aero-2000, International Conference on Human Computer Interaction in Aeronautics*. (pages 123 - 129) Toulouse, France: Cepadues-Editions

Sherry, L., M. Feary, P. Polson, & E. Palmer (2000). Why won't it take that? MCP control device invokes more than one autopilot behavior NASA-Ames Research Center, 2000. Also published as Honeywell Publications C69-5370-14

Abstract: A class of automation surprises can be attributed to the fact that the same mode control panel (MCP) knob results in two distinct autopilot control behaviors depending on the situation. For example, one behavior flies to, and captures, the MCP altitude, the other behavior breaks the capture and flies away from the MCP altitude. This paper describes a formal modeling technique for identifying MCP control devices that change their function in different contexts. This paper also describes how the contents of the formal modeling technique can be used to develop training and to set criteria used for certification.

Sherry, L., M. Feary, P. Polson, & E. Palmer (2000). Formal Method for Identifying Two Types of Automation Surprises. NASA Ames Research Center. Honeywell Publications C69-5370-16.

Abstract: "Automation surprises" occur when operators of sophisticated automation, such as pilots of aircraft, hold a mental model of the behavior of the automation that does not reflect the actual behavior of the automation. This leads to increased workload, and reduced efficiency and safety.

This paper describes a formal method for analysis of automation and its user-interface for two well known characteristics that lead to automation surprises: (1)

an automation user input device that, when selected, results in different automation behaviors depending on the situation, and (2) automation displays that do not provide unique annunciation for all automation behaviors. This method is unique in that it is based on analysis of the goals and behavior of the actual automation software. This provides a meaningful basis to perform user-oriented task analyses. A case study is also provided.

Polson, P. G., & Smith, N. (1999) The cockpit cognitive walkthrough. *Proceedings of the Tenth Symposium on Aviation Psychology*. (pp. 427-432) Columbus OH: Ohio State University.

Abstract: The Cockpit Cognitive Walkthrough is a usability inspection method that evaluates interactions between a cockpit procedure and an avionics interface by showing that the interface supports execution of the procedure by providing feedback for correct pilot actions and error recovery and by guiding the execution of a novel or an infrequently performed procedure. This paper describes the method and summarizes two evaluation studies current in progress.

Smith, N., Moses, J., Polson, P., Romahn, S., Palmer, E. (1999) A Survey of Flight Crew Experience with FANS-1 ATC Datalink. *Proceedings of the Tenth Symposium on Aviation Psychology*. Columbus OH: Ohio State University.

Sherry L., & Polson, P. G. (1999). Shared Models Of FMS Vertical Guidance. *International Journal of Aviation Psychology*, 9, 139-154.

Abstract: Solutions to operational and training issues with flight automation should be based on a complete model of the behaviors of the avionics software that are shared by pilots, developers, and researchers. The operational procedure model, described in this paper, represents the behavior of the avionics software as a set of intentions associated with situation-action rules. This paper focuses on the guidance function incorporated in Flight Management Systems and presents an operational procedure model of vertical guidance.

The source of operational and training issues can be traced to the fact that the operational procedures that determine the behavior of avionics are hidden: (1) they are not trained, (2) they are not annunciated, and (3) they are not known to researchers and designers. We describe research and training programs, based on the model, to uncover, define, and train this hidden logic and functionality.

Feary, M., Polson, P. G., Alkin, M., McCrobie, D., Sherry, L., & Palmer, E. (1998) Aiding Vertical Guidance Understanding. In G. Boy, C. Graeber, & J-M Robert (Eds.) *Proceedings of HCI-Aero-98, International Conference on Human Computer Interaction in Aeronautics* (pp. 165 - 170) Montreal: Ecole Polytechnique de Montreal.

Kitajima, M., Soto, R., & Polson, P. G. (1998) LICAI+: A comprehension-based model of the recall of action sequences. In F. Ritter & R.M. Young (Eds.) *Proceedings of Second European Conference on Cognitive Modelling* . pp. 82-89. Nottingham University Press: Nottingham.

Kitajima, M. & Polson, P.G. (1998). Knowledge Required for Understanding Task-Oriented Instructions. In the *Proceedings of Asia Pacific Computer Human Interaction 1998, APCHI98*, Japan, July 15-17, 1998, pp.19-24. IEEE Computer Society: Los Alamitos, CA.

Sherry, L., & Polson, P. G. (1998). Implications of Situation-Action Rule Descriptions of Avionics Behavior. In G. Boy, C. Graeber, & J-M Robert (Eds.) *Proceedings of HCI-Aero-98, International Conference on Human Computer Interaction in Aeronautics* (pp. 213 - 219) Montreal: Ecole Polytechnique de Montreal.

Sponsored Project Award History

Project Title

Sponsor's Name **National Institute on Deafness and Other Communication Disorders**

Efficacy of Voice Treatment for Parkinson's Disease

Project # **1539651** Speedtype

See Also 1535497, 1539649, 1539650, 1539652

Start **5/17/02** End **3/31/04**

Pending? Project Period Ends **3/31/07**

Animals or Human Subjects?

Proposal No. **H** (A and/or H)

0103.04.0132B
 0301.04.0368B

IDC Rate **FULL IDC?**

26% MTDC

Sponsoring Agency **HHS NIDCD**

Award No. **5 R01 DC001150-12**

Award Total **41,553.00**

Sponsor Type **F**

F = Federal
L = Fed Labs
I = Industry
U = Univ
C = State of Colo
P = Public/Gov
O = Orgs/Assoc
D = Foundations
R = Foreign

Fixed Price? (FP)
ONR (x)
Colo (c)

CFDA# **93.173**

Grant (A-110)
Contract (FAR) **G**
Non-Federal (G, C, or N)

Department **SLHS** College **A&S**

Dept Org # **10216** Box **409**

PIs **Ramig, Lorraine**

PI Email **Lorraine.Ramig@Colorado.EDU**

Phone **492-3023**
FAX

Mod Mo	Mod #	\$ Amount	Modification
10/02	0	21,268.00	
1/03	1	0.00	correct F&A rate/calculation yrs 12-15; see remarks*
1/03	2	0.00	correct yr -11 F&A calculation
3/03	3	20,285.00	extend from 3/31/03 to 3/31/04

Remarks or Pending Award Info

subcontract to Northwestern University
anticipated future funding (includes all associated project numbers):

- 12 yr: \$478,070
- 13 yr: \$621,875
- 14 yr: \$548,390
- 15 yr: \$435,760

*mod 1: "0" added to award no. due to agency computer system going from 5 digits to 6 digits

